



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/421,416	10/19/1999	HINRICH SCHUETZE	D/99198	4124

7590 07/16/2003

JOHN E BECK
XEROX CORPORATION
XEROX SQUARE 20A
ROCHESTER, NY 14644

EXAMINER

FLEURANTIN, JEAN B

ART UNIT	PAPER NUMBER
----------	--------------

2172

DATE MAILED: 07/16/2003

14

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary

Application No.

09/421,416

Applicant(s)

SCHUETZE ET AL.

Examiner

Jean B Fleurantin

Art Unit

2172

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
- Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) ☒ Responsive to communication(s) filed on 29 April 2003.
- 2a) ☐ This action is **FINAL**. 2b) ☒ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) ☒ Claim(s) 7-39, 41-43 and 45-49 is/are pending in the application.
- 4a) Of the above claim(s) _____ is/are withdrawn from consideration.
- 5) ☐ Claim(s) _____ is/are allowed.
- 6) ☒ Claim(s) 7-22, 28, 39, 41-43, 45 and 47 is/are rejected.
- 7) ☒ Claim(s) 23-27, 29-38, 46 and 48 is/are objected to.
- 8) ☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☐ The drawing(s) filed on _____ is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.
- Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
- 11) ☐ The proposed drawing correction filed on _____ is: a) ☐ approved b) ☐ disapproved by the Examiner.
- If approved, corrected drawings are required in reply to this Office action.
- 12) ☐ The oath or declaration is objected to by the Examiner.

Priority under 35 U.S.C. §§ 119 and 120

- 13) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some * c) ☐ None of:
1. ☐ Certified copies of the priority documents have been received.
 2. ☐ Certified copies of the priority documents have been received in Application No. _____.
 3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).
- * See the attached detailed Office action for a list of the certified copies not received.
- 14) ☐ Acknowledgment is made of a claim for domestic priority under 35 U.S.C. § 119(e) (to a provisional application).
- a) ☐ The translation of the foreign language provisional application has been received.
- 15) ☐ Acknowledgment is made of a claim for domestic priority under 35 U.S.C. §§ 120 and/or 121.

Attachment(s)

- | | |
|--|---|
| 1) <input checked="" type="checkbox"/> Notice of References Cited (PTO-892) | 4) <input type="checkbox"/> Interview Summary (PTO-413) Paper No(s). _____ |
| 2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948) | 5) <input type="checkbox"/> Notice of Informal Patent Application (PTO-152) |
| 3) <input type="checkbox"/> Information Disclosure Statement(s) (PTO-1449) Paper No(s) _____ | 6) <input type="checkbox"/> Other: _____ |

DETAILED ACTION

Response to Amendment

1. Claims 7-39, 41-43 and 45-49 are remained pending for examination.

Response to Arguments

2. Applicant's arguments, see Tanaka does not teach what it is alleged to teach. In particular, in contrast to all of the amend independent claims, Tanaka does not teach a method for processing digital documents, filed on 4/29/03, with respect to the rejection(s) of claim(s) 7-22, 28, 39, 41-43, 45, 47 and 49 under 35 U.S.C. 103(a) have been fully considered and are persuasive. Therefore, the rejection has been withdrawn. However, upon further consideration, a new ground(s) of rejection is made in view of Douglass R. Cutting et al. ("Douglass") as indicated in section 3.

Claim Rejections - 35 U.S.C. § 103

3. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

Art Unit: 2172

Claims 7-22, 28, 39, 41-43, 45, 47 and 49 are rejected under 35 U.S.C. 103(a) as being unpatentable over “A Cluster-based Approach to Browsing Large Document Collections, 1992” issued to Douglas et al. (‘Douglas’).

As per claims 1 and 39, Douglas teaches a method for quantitatively representing documents in a vector space, as claimed comprises the steps of identifying a first document to be processed from a plurality of objects documents (thus, numerous document similarity measures have been proposed all of which treat each document as a set of words, often with frequency information and measure the degree of word overlap between documents, the documents are typically represented by sparse vectors length equal to the number of unique words in the corpus, each component of the vector has a value reflecting the occurrence of the corresponding word in the document; which is readable as identifying a first document to be processed from a plurality of objects documents)(see pages 320 and 321, cols. 2 and 1, lines 25-26 and 1-7);

extracting a first feature corresponding to the first document from the plurality of documents, the first feature comprising text surrounding an image included in the document, the text surrounding the image not being anchor text (thus, the browsing component describes groups of similar documents, one or more of which can be selected for further examination, this can be iterated until the user is directly viewing individual documents; which is readable as extracting a first feature corresponding to the first document from the plurality of documents, the first feature comprising text surrounding an image included in the document, the text surrounding the image not being anchor text; which is readable as extracting a first feature corresponding to the first document from the plurality of documents, the first feature comprising text surrounding an image included in the document, the text surrounding the image not being

Art Unit: 2172

anchor text). But, Douglass does not explicitly indicate converting the first feature to a first vector; and associating the first vector with the first document. However, Douglass implicitly indicates for each document α in a collection C , let the countfile $c(\alpha)$ be the set of words with their frequencies that occur in that document, let V be the set of unique words occurring in C , then $c(\alpha)$ can be represented as a vector of length absolute value of V ; which is readable as converting the first feature to a first vector, (see page 322, col. 1, lines 1-8); and each component of the vector has a value reflecting the occurrence of the corresponding word in the document; which is readable as associating the first vector with the first document, (see page 321, col. 1, lines 5-7). It would have been obvious to a person of ordinary skill in the art at the time the invention was made to modify the teachings of Douglass with converting the first feature to a first vector; and associating the first vector with the first document. This modification would allow the teachings of Douglass to improve the accuracy and the reliability of the system and method for representing data objects in vector space, and provide a powerful new access paradigm, (see page 318, col. 1, line 11).

As per claims 7, 10 and 15, Douglass teaches a method as claimed, further comprises the steps of extracting a second feature corresponding to the document (thus, text search methods such as near neighbor search or snippet search, the browsing component describes groups of similar documents one or more of which can be selected for future examination; which is readable as extracting a second feature corresponding to the document)(see page 319, col. 1, lines 30-40). But, Douglass does not explicitly indicate converting the second feature to a second vector; and associating the second vector with the first document. However, Douglass implicitly indicates for each document α in a collection C , let the countfile $c(\alpha)$ be the set

Art Unit: 2172

of words with their frequencies that occur in that document, let V be the set of unique words occurring in C , then $c(\alpha)$ can be represented as a vector of length absolute value of V ; which is readable as converting the second feature to a second vector, (see page 322, col. 1, lines 1-8); and each component of the vector has a value reflecting the occurrence of the corresponding word in the document; which is readable as associating the second vector with the first digital document, (see page 321, col. 1, lines 5-7). It would have been obvious to a person of ordinary skill in the art at the time the invention was made to modify the teachings of Douglass with converting the second feature to a second vector; and associating the second vector with the first document. This modification would allow the teachings of Douglass to improve the accuracy and the reliability of the system and method for representing data objects in vector space, and provide a powerful new access paradigm, (see page 318, col. 1, line 11).

As per claims 8, 11, 13, 16, 18, 19 and 41, the limitations of claims 8, 11, 13, 16, 18, 19 and 41 are rejected in the analysis of claim 7, and these are rejected on that basis.

As per claims 9 and 14, the limitations of claims 9 and 14 are rejected in the analysis of claim 7, and these are rejected on that basis.

As per claims 12, 17 and 42, Douglass teaches the method as claimed, wherein the numeric value representative of the number of links in each corresponding document linking to the document is calculated as the token frequency weight of the corresponding link multiplied by the inverse context frequency weight of the corresponding link (thus, a document group is in some sense dual to the trimmed sum profile, rather than consider the central documents of a cluster, we define $tw(T)$, to be w highest weighted terms; which is readable as wherein the numeric value representative of the number of links in each corresponding document linking to

the document is calculated as the token frequency weight of the corresponding link multiplied by the inverse context frequency weight of the corresponding link)(see page 322, col. 2, lines 14-20).

As per claim 20, the limitations of claim 20 are rejected in the analysis of claim 7, and this claim is rejected on that basis.

As per claim 21, in addition to the discussion in claim 1 and 7, Douglass further teaches for each possible text genre, processing the first to calculate the probability that the first document is of the corresponding text genre (thus, compute the probability that if we choose a sample of size s_1 we fail to get any individual from some cluster, this is at most k times the probability that none of our s individual is a member of cluster i namely; which is readable as , processing the first to calculate the probability that the first document is of the corresponding text genre)(see page 325, col. 1, lines 20-26).

As per claims 22, 28, 45 and 47, Douglass teaches a method as claimed, wherein the first feature comprises the color histogram for the image included in the first document (thus, numerous document similarity measures have been proposed all of which treat each document as a set of words, often with frequency information and measure the degree of word overlap between documents, the documents are typically represented by sparse vectors length equal to the number of unique words in the corpus, each component of the vector has a value reflecting the occurrence of the corresponding word in the document; which is readable as wherein the first feature comprises the color histogram for the image included in the first document)(see pages 320 and 321, cols. 2 and 1, lines 25-26 and 1-7).

As per claims 43 and 49, in addition to the discussion in claims 1 and 7, Douglass further teaches selecting an image feature as a first feature, the image feature being associated with the non-text content of an image included in the document (thus, text search methods such as near neighbor search or snippet search, the browsing component describes groups of similar documents one or more of which can be selected for further examination; which is readable as selecting an image feature as a first feature, the image feature being associated with the non-text content of an image included in the document)(see page 319, col. 1, lines 36-43).

As per claim 44, the limitations of claim 44 are rejected in the analysis of claim 22, and this is rejected on that basis.

Allowable Subject Matter

4. Claims 23-27, 29-38, 46 and 48 are objected to as being dependent upon a rejected base claim, but would be allowable if rewritten in independent form including all of the limitations of the base claim and any intervening claims.

Prior Art

5. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure. Wong et al "Generalized Vector Space Model In Information Retrieval," 1985.

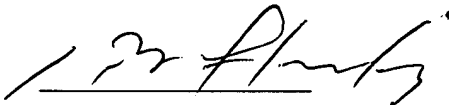
Art Unit: 2172

Contact Information

6. Any inquiry concerning this communication from examiner should be directed to Jean Bolte Fleurantin at (703) 308-6718. The examiner can normally be reached on Monday through Friday from 7:30 A.M. to 6:00 P.M.

If any attempt to reach the examiner by telephone is unsuccessful, the examiner's supervisor, Mrs. KIM VU can be reached at (703) 305-8449. The FAX phone numbers for the Group 2100 Customer Service Center are: *After Final* (703) 746-7238, *Official* (703) 746-7239, and *Non-Official* (703) 746-7240. NOTE: Documents transmitted by facsimile will be entered as official documents on the file wrapper unless clearly marked "**DRAFT**".

Any inquiry of a general nature or relating to the status of this application or proceeding should be directed to the Group 2100 Customer Service Center receptionist whose telephone numbers are (703) 306-5631, (703) 306-5632, (703) 306-5633.



Jean Bolte Fleurantin

07/9/03

JBF/



JEAN M. CORRIELUS
PRIMARY EXAMINER